

Unzipping DNA : A hypothesis on correlation during replication

Somendra M Bhattacharjee

Institute of Physics, Bhubaneswar-751 005,
Orissa, India

E-mail : somen@iopb.res.in

Abstract Based on the result of an unzipping phase transition by a force in a model of DNA, we hypothesize that the DnaA-type proteins act as a pulling agent with a force slightly less than the critical force for unzipping. The dynamic (space-time) correlation of unzipping then drives the subsequent events of replication. In such a correlation driven scenario, there is no need of a replisome as a structural unit and this elusive replisome may not exist at all.

Keywords : DNA, unzipping phase transition, biomolecules theory and modeling

ACS Nos. 87.14 Gg, 87.15 He, 64.60.Cn

1. Introduction

A double stranded DNA (dsDNA) needs to be opened up for its replication [1]. Once done, the subsequent processes take place sequentially, like the binding of, say, helicases, single-strand-binding proteins *etc*, then the elongation of the new chain by polymerases *etc*.

Recently, it has been proposed [2] that the action of DnaA-type proteins could be thought of as a force acting at the initiation site (called 'origin'). (The function of DnaA is not to be confused with the 'locomotive' or 'sewing machine' action of the hypothetical replisome.) A simple model showed that there is a phase transition at a critical value of the force so that for forces less than the critical strength, DNA is a double-stranded object while for forces exceeding it, the chains get opened up.

In this paper, we give a short description of this model, the principal results and the subsequent developments. We point out the features that seem to have led to 'confusions'. We also show how this simple model can be extended to incorporate some other features.

The connection to *in vivo* replication and laboratory experiments are discussed in section 3, where we propose that the correlation of unzipping is at the heart of the highly correlated collective process that goes by the name of replication.

2. Statistical mechanical model

A. Model and results

The model proposed originally is that of a pair of flexible polymers bound together by a short-range interaction and pulled at one end by a force [2]. We start with a homo-DNA, *i.e.* all base pairs identical. One end of the two strands are anchored or tied together and the other end is pulled by a force g . The Hamiltonian is given by

$$\frac{H}{k_B T} = \int_0^N ds \left[\frac{\epsilon_1}{2} \left(\frac{\partial \mathbf{r}_1(s)}{\partial s} \right)^2 + \epsilon_2 \left(\frac{\partial \mathbf{r}_2(s)}{\partial s} \right)^2 + V(\mathbf{r}(s)) \right] + H_{force} \quad (1a)$$

$$V(\mathbf{r}) = v \delta(\mathbf{r}) \quad (1b)$$

$$H_{force} = -g \cdot (\mathbf{r}(N)) = -g \cdot \int_0^N ds \frac{\partial \mathbf{r}(s)}{\partial s}, \quad (1c)$$

where $\mathbf{r}_i(s)$ denotes the d -dimensional position of a monomer on chain i (of elastic constant ϵ_i) at a contour-length s measured from the tied end ($s = 0$), $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is the relative separation of the two chains at the same monomer s and $V(\mathbf{r})$ is a short-range potential simulating the interaction of the base pairs of two strands. Note that $\mathbf{r} = 0$. Any realistic potential can be chosen

here. But our interest is in the effect of the force on the bound double-strand case, and so for analytical simplicity, we choose a delta function (or contact) potential, to be interpreted as a limit of a narrow square well. The strength of the well is chosen to be such that there is only one bound state in the problem.

As usual in statistical mechanics, there are two possible ensembles, namely a constant force (g constant) ensemble and a constant end-point separation ($r(N) = \text{constant}$) ensemble. For studies of phase transitions, it is useful to work in an ensemble of fixed intensive variable. We therefore, choose the fixed force ensemble.

For the zero force case, there is a critical unbinding transition at $v = v_c$. The pulling force would like to align the strands in the direction of the force while the interaction would like to keep the strands together. There is a critical force below which the DNA remains in the ds-DNA phase while for forces exceeding this critical value it gets unzipped.

Quantitatively, the partition function can be evaluated by a transfer matrix approach with the contour length as the transfer direction. The free energy per unit length is given by the largest eigenvalue of the transfer matrix and a phase transition takes place whenever there is a degeneracy of the largest eigenvalue.

The transfer matrix can conveniently be written in the form of a quantum Hamiltonian, albeit non-hermitian [3], for a particle of coordinate r (CM behaviour is like a free chain or a free particle)

$$H_q(g) = \frac{1}{2}(p + ig)^2 + V(r). \quad (2)$$

in units of $\hbar (\equiv k_B T) = 1$ and mass = 1, with p as momentum. For long chains $N \rightarrow \infty$, the free energy is the ground-state energy of this non-hermitian Hamiltonian. A phase transition takes place whenever the ground state is degenerate. The analysis done in Ref. [2] shows that if the ground state energy (i.e. the binding energy of the dsDNA per unit length or per base pair) is E_0 , then the critical force is given by

$$g_c = 2\sqrt{E_0} - |v - v_c|^{1/2-d} \quad (3)$$

where the v -dependences [4] of E_0 close to v_c , for general d , is used. In absence of any force, DNA can be denatured either by changing temperature or by changing pH of the solution. In this

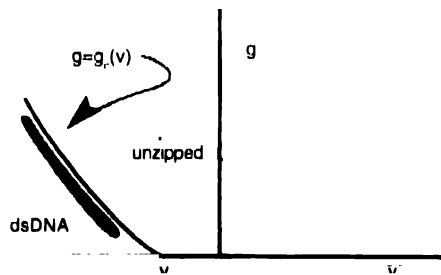


Figure 1. The $g-v$ phase diagram. Thermal or pH driven denaturation without any force is the $g = 0$ line. The DnaA protein is hypothesized to work in the shaded region just below the critical line.

respect, the effective interaction parameter v is a better variable than the temperature itself. The generic phase diagram is shown in Figure 1.

The degeneracy of the eigenvalues leads to a diverging length scale that could be associated for $g \rightarrow g_c$, as the characteristic length of the opened Y-fork (the replication fork). This length scale measured along the strand of the DNA, gives a measure of the number of monomers opened by the force. Denoting this number of m , we get

$$\langle m \rangle \sim |g_c - g|^{-1} \quad \text{with} \quad \frac{1}{2} = 1. \quad (4)$$

(v_m was denoted by v_r in Ref [2].)

B. Questions on the model and the results

It is known for centuries that a certain minimum pulling force is needed to move an object on a surface with friction, or for adhesion, and so, is the occurrence of a critical force an obvious result? An understanding of this threshold phenomena beyond phenomenological level is still in its infancy. Our results are to be seen in this context, and the model is different from the stick-slip models used in friction or adhesion problems. In fact previous force measurements [5] for unzipping of DNA have been interpreted in terms of equilibrium statistical mechanics and not as a friction problem.

A classical diatomic molecule or a bond (or a particle in a potential well) under a force, does not show any phase transition. At a quantum mechanical level, a bond with a stretching force is still described by a hermitian Hamiltonian and therefore, will not show the phase transition of the previous subsection. It is the polymeric property or connectivity that is crucial for the unzipping transition.

The result of eq. (4) is not to be confused with the pulling of a polymer discussed in text books. In the case of a polymer pulled at the ends by a stretching force, there would be an extension of the chain and for a small force, the overall dimension of the polymer would scale linearly with the stretching force [6]. In contrast, we see that, in the dsDNA situation, a small force will have no major effect and the law in eq. (4) is a dependence on the *deviation from the critical force*.

Is continuum model a good approach? The answer to this question lies in the recognition of various length scales. The Hamiltonian of eq. (1a) can be discretized and studied with appropriate bond lengths and realistic $V(r)$. We adopted a coarse-grained approach to focus on the phase transition behaviour. In other words, a continuum approach by itself is not a necessity to study the phase transition and other details of unzipping.

Previous studies seem to indicate that DNA is better represented by worm-like chains but again a renormalization group argument shows that in the long length limit such a Hamiltonian will generate the elastic term [7] and therefore eq

(1a) is a representative of the universality class. In the same spirit, the delta-function potential can be replaced by any realistic potential. But for the binding-unbinding (melting) transition, it is only the integral of the potential that matters and a square well or delta function is a valid starting point.

Is the restriction to homo-DNA any good ? It is worth remembering the role played by homo-DNA's in deciphering the genetic code. For a controlled experiment to detect the unzipping transition, large homo-DNA's should be studied. The model is easily generalizable (see below) to consider specific base-sequence [8].

Subsequent developments

After the original proposal of the model for unzipping in Ref. [2], several studies have been made. The dynamics of pulling a polymer from a potential well, *i.e.*, the dynamics of unzipping in one dimension has been studied at a mean field level by Sebastian [9]. In addition to recovering the results of previous section, Ref [9] shows that the unzipping transition is also characterized by a diverging time scale. Denoting the characteristic time scale τ , one finds

$$\tau \sim g_c - g \quad \text{with} \quad \frac{d\tau}{dg} = 1. \quad (5)$$

A divergence of time scales indicates a long-range correlation in time in dynamics of unzipping.

DNA generally consists of inhomogeneous sequence of base pairs. Such a case in this model can be considered by making the interaction energy to be monomer position dependent. The interaction term in eq. (1a) can be written as $\int ds b(s) v \delta(\mathbf{r})$, with $b(s)$ a variable depending only on monomer (base) position s describing the specific details of the sequence. A case of random $b(s)$, a random interaction model (RANI model) has been considered earlier [9] for the zero force case where the randomness has been found to be marginal. In the case of unzipping transition, the randomness does not destroy the sharpness of the transition but changes [10] the exponent ν_m defined in eq. (4) to $\nu_m = 2$. This is for an ensemble-averaged single molecule. The dynamics is yet to be studied.

The nature of the transition has also been considered. Maritan, Orlandini and Seno [11] showed by Monte Carlo simulations and analytical methods that the noncrossing constraint on the chains, can lead to a first order transition, and more interestingly to a 'cold denaturation'. It has recently been argued [12] from a direct evaluation of the partition function that the transition for the Hamiltonian of eq. (1a) for $d=1$ could be first order. It is known that depending on the reunion exponent, the phase transition for polymers could be first order [13-16], and in the quantum context, the order of transition (eq. [2]) is determined by the normalizability of the 'critical' wavefunction [17, 18]. Several exact results for the unzipping transition in various dimensions have been obtained in Ref. [19].

On the experimental side, atomic force microscopes [20,21] have been used to measure the force vs extension curves for DNA. Such experiments have revealed a reversible unzipping transition, but with hysteresis, indicating a first-order transition. However a detailed quantitative study of the unzipping is still awaited.

D. Extensions

The simple model does not take into account the self avoidance of the chains. The effects of self-avoidance have been studied in Ref. [11]. In a laboratory experiment, one can take eq. (1a) as a model for DNA in a theta-solvent [6]. In a good solvent, the self and mutual avoidance of the chains can be introduced by an imaginary random potential and averaging the partition function [22]. Adding a term to eq. (1a)

$$\frac{H_{imag}}{k_B T} = \int_0^N ds [iV(\mathbf{r}_1(s)) + iV(\mathbf{r}_2(s))]. \quad (6)$$

with $V(\mathbf{r})$ as an annealed Gaussian random potential with zero mean and variance $\langle V(\mathbf{r}) V(\mathbf{r}') \rangle = u \delta(\mathbf{r} - \mathbf{r}')$, the averaged partition function $\bar{Z} = \int D\mathbf{R} \exp(-H)$, with overline denoting averaging over V , leads to a hamiltonian of the form of eq. (1a) with additional terms, H_{int} ,

$$H_{int} = \int ds ds' \sum_{i,j=1,2} u \delta(\mathbf{r}_i(s) - \mathbf{r}_j(s')). \quad (7)$$

This represents excluded volume interaction for the chains. This imaginary potential of eq. (6) would lead to a complex scalar potential in the quantum Hamiltonian of eq. (2).

We have ignored the double helical nature of DNA. In actual replication, topoisomerases act when there is super-coiling. Once the chains are opened up in a region by DnaA forming a Y-fork, the remaining chain is brought back to the native state by the topoisomerase. This gives a justification for the absence of the helical configuration in the hamiltonian in eq. (1a). The topological linking number of the two chains can be described by the Gauss integral which is known to act like a real vector potential [22]. It would be appropriate to take into account the linking as an independent quantity only if there is a change in the coiling in the process [23].

A toy model is a quantum Hamiltonian corresponding to a polymer with winding as

$$H_q = \frac{1}{2} (p + A)^2 + V(\mathbf{r}) + iV(\mathbf{r}), \quad (8)$$

where $A = A_0(\mathbf{r}) + i g$, with the real vector potential describing windings and the imaginary part, as before, the exerted force. Studies of the most general quantum Hamiltonian with both complex scalar and vector potential are needed.

3. Implication for biological replication

We explain the hypothesis on the role of unzipping in replication.

The replication process starts with a DnaA protein attaching itself, in interaction with the membrane, at the 'origin' to start the Y-fork (or 'eye'). The next step is the binding of the various enzymes/proteins like single-strand binding (SSB) protein, helicase, topoisomerase, polymerases *etc.* Unlike the latter proteins/enzymes, the functionality of DnaA protein is not well understood [24]. It is apparent that the whole process requires a strong correlation in space and time, whose source or origin is not obvious. In order to explain such correlations in subsequent processes, a structural unit 'replisome' has been postulated which has never been isolated.

We make the hypothesis that

- The function of DnaA is actually to exert a pulling force (let us take this time $t = 0$) with a force which is close to, but slightly below, the critical force for unzipping. This force tends to form a Y-fork or an eye-type bubble (depending on the location of the origin).
- The large length and time scales for unzipping process, leads to a space time correlation in the initial nonequilibrium dynamics of unzipping. We postulate that this correlation controls the subsequent processes especially the dynamics of bindings of the subsequent enzymes/proteins.

The correlation in the unzipping dynamics can be characterized quantitatively by the early time behaviour of, say, $\langle m(t)m(t' \rightarrow 0) \rangle$ where $m(t)$ is the number of monomers unzipped at time t , or a more microscopic correlation $\langle r_i(s, t)r_j(s' \rightarrow 0) \rangle$, i, j being the spatial components.

The dynamics of unzipping is described by a Langevin equation

$$m \frac{\partial r(s, t)}{\partial t} = - \frac{\partial H}{\partial r(s, t)} + \eta(s, t), \quad (9)$$

where t represents time, H is given by eq. (1a), η is the thermal noise related to friction coefficient μ by $\langle \eta_i(s, t)\eta_j(s', t') \rangle = 2\delta_{ij}\mu k_B T \delta(s-s') \delta(t-t')$. The binding of the next set of proteins in the opened region may be represented symbolically by a kinetic equation of the type

$$\frac{\partial c}{\partial t} = -C(s, t) f\{c\}, \quad (10)$$

where c denotes the concentration of the free enzymes in the bath, $f\{c\}$ describes the dynamics in the absence of any correlations, and $C(s, t)$ stands for the initial or early time

behaviour of the unzipping correlations. This is equivalent to making the rate constant dependent on the initial correlation [25]. This way of viewing the replication process as a coupled system driven by the long-range space-time correlated unzipping dynamics is different from the prevalent contact based approach.

It is necessary to (a) study long homo-DNA's under well controlled pulling force to map out the phase diagram [2, 11] and (b) probe jointly the correlation and binding kinetics in the presence of a pulling force.

A few appealing features of such a scenario are to be noted. Since the critical force is dependent on the structural details of the DNA, which is different for different species, our hypothesis naturally requires species variations of the DnaA-type protein for proper functionality. This protein should be such that it exerts a force slightly below the critical value for that particular DNA. Finally, in the proposed scenario, the dynamic correlation plays the most important role, and therefore a composite structural unit is not a necessity. In other words, the so-far-elusive hypothesized replisome need not exist as a structural unit.

References

- [1] A Kornberg and T Baker *DNA Replication* 3rd edn (San Francisco: W. H. Freeman) (1992)
- [2] S M Bhattacharjee *cond-mat/9912297* (available from <http://xxx.lanl.gov> or its mirror) *J. Phys.* **A33** L423 (2000)
- [3] N Hatano and D R Nelson *Phys. Rev.* **B56** 8651 (1997)
- [4] E B Kolomeisky and J P Straley *Phys. Rev.* **B46** 1260 (1992)
- [5] U Bockelmann, B Essevez-Roulet and F Heslot *Phys. Rev. E* **5** 2386 (1998)
- [6] P G de Gennes *Scaling Concepts in Polymer Physics* (Ithaca: Cornell University Press) (1979)
- [7] K Freed *Renormalization Group Theory of Macro-molecules* (New York: John Wiley) (1987)
- [8] S M Bhattacharjee and S Mukherji *Phys. Rev. Lett.* **70** 49 (1993) **70** 3359(E) (1993); *Phys. Rev.* **E48** 3483 (1993)
- [9] K L Sebastian *Phys. Rev.* **E62** 1128 (2000)
- [10] D Lubensky and D R Nelson *Phys. Rev. Lett.* **85** 1572 (2000)
- [11] A Maritan, E Orlandini and F Seno University of Padova preprint (2000)
- [12] Haijun Zhou *cond-mat/0007015* (available from <http://xxx.lanl.gov> or its mirror)
- [13] M E Fisher *J. Stat. Phys.* **34** 667 (1984)
- [14] S M Bhattacharjee and S Mukherji *Phys. Rev. Lett.* **83** 217 (1999)
- [15] Y Kafri, D Mukamel and L Peliti *Phys. Rev. Lett.* **85** 4988 (2000) (*cond-mat/0007141*)
- [16] M S Causo, B Coluzzi and P Grassberger *Phys. Rev.* **E62** 395 (2000)
- [17] S Mukherji and S M Bhattacharjee *cond-mat/0101128*, R K Zia, R Lipowsky and D M Kroll *Am. J. Phys.* **56** 160 (1988)

- [18] Another well-known example of criticality with first-order transition is the ferroelectric six vertex model J F Nagle *Comm. Math. Phys.* **13** 62 (1969); R J Baxter *Exactly Solved Models in Statistical Physics* (New York : Academic) (1982)
- [19] D Marenduzzo, A Trovato and A Maritan *Phys. Rev.* **E64** 031901 (2001)
- [20] M Reif, H Clausen-Schaumann, H E Gaub *Nature Struc. Biol.* **6** 346 (1999)
- [21] D Anselmetti, J Fritz, B Smith and X Fernandez-Busquets *Single Mol.* **1** 53 (2000)
- [22] S F Edwards in *Polymer Physics : 25 Years of the Edwards Hamiltonian* (ed) S M Bhattacharjee (Singapore : World Scientific) (1992); M Doi and S F Edwards *Dynamics of Polymer Solutions* (Oxford . Oxford U Press) (1986)
- [23] J F Marko and E D Siggia *Phys. Rev.* **E52** 2912 (1995)
- [24] See e.g. Plate 15 of Ref [1]
- [25] R Zwanzig *J. Chem. Phys.* **97** 3587 (1992)